

REMARKS

In the Office Action, the Examiner rejected Claims 1-11, 13-15, 19-23 and 27-30, which are all of the pending claims, under 35 U.S.C. 103 as being unpatentable over a combination of four references. Specifically, these claims were rejected as being unpatentable over U.S. Patent 6,405,161 (Goldsmith) in view of U.S. Patent 7,103,536 (Kanno) further in view of U.S. Patent 6,098,034 (Rizin, et al.) and further in view of U.S. Patent 7,167,825 (Potter). Claims 21-23, 27 and 28 were also rejected under 35 U.S.C. 101 as directed to non-statutory subject matter.

The previous rejection of the claims as being unpatentable based on the combination of Goldsmith and Kanno was withdrawn

Independent Claims 1, 13 and 21 are being amended to better define the subject matters of these claims. In particular, these claims are being amended to describe in more detail the function of the Patricia tree that is described in these claims.

The rejection of Claims 21-23, 27 and 28 under 35 U.S.C. 101 is respectfully traversed.

In rejecting these claims as non-statutory, the Examiner, in the Office Action, argued that these claims are directed to a computer readable medium and are not limited to a non-transitory, and thus, statutory medium. The Examiner further argued that the scope of “program storage device” can include non-statutory instances.

Applicants respectfully disagree because the program storage device defined by Claim 21 is expressly limited to “tangibly embodying a program of instructions executable by a machine. Thus, non-tangible, non-statutory data storage does not come within the scope of the claim. In addition, in view of the requirement that the data is stored in the medium in a tangible way, the medium is an apparatus or an article of manufacture that comes within the scope of patent eligible subject matter under U.S.C. 101.

In order to emphasize the statutory nature of claims 15-19, Claim 15 is being amended to include expressly the limitations that “The program storage device” is a tangible program storage device. With this limitation, it is clear that the claim is directed to a physical, tangible article of manufacture or apparatus, and is patent eligible within the meaning of 35 U.S.C. 101. The Examiner is, accordingly, requested to reconsider and to withdraw the rejection of Claims 15-19 under 35 U.S.C. 101.

Moreover, for the reasons discussed below, Claims 1-11, 13-15, 19-23 and 27-30 patentably distinguish over the prior art and are allowable. The Examiner is thus also asked to reconsider and to withdraw the rejection of Claims 1-11, 13-15, 19-23 and 27-30 under 35 U.S.C. 103, and to allow these claims.

Generally, Claims 1-15, 19-23 and 27-29 patentably distinguish over the prior art because the prior art does not show or suggest forming and using Patricia trees, as described in independent Claims 1, 13 and 21, to identify sets of candidate prefixes and candidate suffixes.

In order to best understand this feature and its significance, it may be helpful to review briefly the present invention and the prior art.

Applicants’ invention, generally, relates to automatically collecting affixes of a language from one or more documents. As discussed in detail in the present application, knowledge of affixes is important for analyzing existing words and for producing new words. It is very difficult and time extensive to acquire a complete list of affixes of a language by hand, and a number of procedures have been tried to use a computer or some other automated process to identify affixes.

The previous approaches tend to parse words into pieces, either a prefix and a stem, or a stem and a suffix. Also, the prior art approaches tend to limit the length of an affix to reduce the size of the search space.

The prior art systems have a number of disadvantages and limitations. For example, they may fail to discover both prefixes and suffixes at the same time, and they may not be able to discover nested affixes. In addition, because of the limitations on length, the prior art systems may not find many affixes that appear in technical documents. Further, many of the previous approaches fail to find affixes containing non-alphabet characters such as digits and hyphens.

The present invention effectively addresses these prior art limitations. Generally, the present invention provides an unsupervised, knowledge-free procedure for automatically discovering prefixes and suffixes from text. The present invention integrates prefix and suffix discovery in such a way that uses knowledge about prefixes to find suffixes and uses knowledge about suffixes to find prefixes.

More specifically, in one embodiment, the invention provides a computer system for analyzing text in one or more documents, and comprising one or more system interfaces, and an affix process that determines one or more affixes of one or more words in one or more of the documents and provides the affixes to the system interface. This affix determining process comprises obtaining a collection of words, and representing all of the words in the collection as Patricia trees to show visually morphological structures of the words to facilitate identifying potential candidates of prefixes and suffixes from the words in the collection.

To form these Patricia trees, the words in the collection are first used to construct first and second tries. Each of these tries has a multitude of paths and a multitude of nodes, and these tries are then compressed by compressing all unary paths in the tries to form the prefix Patricia

tree and the suffix Patricia tree. In this way, words are added into the Prefix tree, and the Prefix tree is used to identify a set of candidate prefixes. Also, in the formation of the suffix Patricia tree, the words in the prefix Patricia tree are reversed, and the reversed words are added into the suffix Patricia tree. This suffix Patricia tree is then used to identify a set of candidate suffixes.

These sets of candidate prefixes and suffixes are then refined, using knowledge previously discovered, to identify actual prefixes and suffixes. In particular, knowledge of prefixes previously identified in the refining is used to further refine the set of candidate suffixes, and knowledge of suffixes previously identified in the refining is used to further refine the set of candidate prefixes.

The references of record do not disclose or render obvious forming and using Patricia trees in the above-described manner to identify sets of candidate prefixes and sets of candidate suffixes from the words in the given collection of words.

Goldsmith discloses an automated, morphological analysis of a natural language for determining prefixes, suffixes and stems. This analysis has three major components: determining the correct morphological split for individual words, establishing accurate categories of stems based on the range of suffixes that they accept, and identifying allomorphs of the same stem.

In the procedure described in Goldsmith, word stems and suffixes are identified and combined as signatures, and then prefixes are identified in the list of stems and suffixes. As the Examiner noted in the Office Action, column 5, lines 51-61, of Goldsmith discloses that prefixes are identified by removing common sequences of one or more letters from the above-mentioned signatures.

Kanno describes a process for identifying a suffix, and the Examiner cited this reference for the disclosure of reversing a word and adding it to a Patricia tree for suffix matching. Applicants' process involves much more than reversing words and adding those reversed words to a Patricia tree.

There are a number of important differences between Applicants' claimed invention and the procedures described in Goldsmith and Kanno. One significant difference is that do not disclose the use of Patricia trees to identify candidate sets of prefixes and suffixes from the words in the given collection. In particular, Kanno discloses matching prefixes, but not identifying candidate prefixes.

With the present invention, the Patricia trees are formed by initially forming first and second tries and then compressing these tries to form the prefix Patricia tree and the suffix Patricia tree. In particular, in Applicants' invention, each of the tries has a multitude of paths and a multitude of nodes, and the tries are compressed by compressing all unary paths in the tries to form the prefix and suffix Patricia trees. In this process, words are added into the prefix Patricia tree; and, in the formation of the suffix Patricia tree, the words in the prefix Patricia tree are reversed, and the reversed words are added into the suffix Patricia tree. The Prefix tree is used to identify a set of candidate prefixes, and the suffix Patricia tree is used to identify a set of candidate suffixes.

The other references of record have been reviewed, and these other references, whether considered individually or in combination, also do not disclose or suggest forming and using Patricia trees as they are formed and used in the present invention.

For example, Rizin, et al. describes a method and system for standardizing phrases in a document. In the disclosed process, a preliminary list of standard phrases is created, and that

preliminary list is filtered to create a final list of standard phrases. A phrase substitution is computed to determine the appropriate conformation of standard phrases to approximate phrases or the approximate phrase to the standard.

Potter discloses a procedure for linguistically analyzing a text to produce text components and for determining a plurality of formulations alternatives for the text.

Potter was specifically cited by the Examiner for representing all of the words in a collection as Patricia trees to show visually morphological structures of the words.

It is important, however, that Potter does not use this Patricia tree as the Patricia tree is used in the present invention – to facilitate identifying potential candidates of prefixes and suffixes from a collection of words.

Independent Claims 1, 13 and 21 are being amended to emphasize the above-described aspect of the present invention. In particular, each of Claims 1, 13 and 21 describes the way in which the Patricia tree. Claims 1, 13 and 21 are being amended to elaborate on the way the Patricia trees are used. In particular, these claims are being amended to include positively the limitation that the Patricia tree represent all of the words in the collection “to show visually morphological structures of the words to facilitate identifying potential candidates of prefixes and suffixes from the words in the collection.”

In light of the above-discussed differences between Claims 1, 13 and 21 and the prior art, and because of the advantages associated with those differences, Claims 1, 13 and 21 patentably distinguish over the prior art and are allowable. Claims 2-11 are dependent from Claim 1 and are allowable therewith; and Claims 14, 15, 19, 20, 29 and 30 are dependent from Claim 13 and are allowable therewith. Likewise, Claims 22, 23, 27 and 28 are dependent from, and are allowable with, Claim 21. The Examiner is, accordingly, asked to reconsider and to withdraw the rejection

of Claims 1-11, 13-15, 19-23 and 27-29 under 35 U.S.C. 103, and to allow these claims and new Claim 30.

For the reasons discussed above, the Examiner is requested to reconsider and to withdraw the rejection of Claims 221-23, 27 and 28 under 35 U.S.C. 101; and the Examiner is asked to reconsider and to withdraw the rejection of Claims 1-11, 13-15, 19-23 and 27-29 under 35 U.S.C. 103, and to allow these claims and new Claim 30.

If the Examiner believes that a telephone conference with Applicants' Attorneys would be advantageous to the disposition of this case, the Examiner is asked to telephone the undersigned.

Respectfully Submitted,

/John S. Sensny/

John S. Sensny
Registration No. 28,757
Attorney for Applicant

Scully, Scott, Murphy & Presser, P.C.
400 Garden City Plaza, Suite 300
Garden City, New York 11530
(516) 742-4343

JSS:gc